Vol. 8, No. 2, April 2026

Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

A Hybrid Certainty Factor–XGBoost Approach for Cyberattack Detection Using the TON_IoT Dataset

Adiva Dwi Aprianto[1], Ratih Hafsarah Maharrani[2*], Indi Cahya Ratna Auliya[3], Vania Rizky Alifiah[4]

[1,2,3,4]Department of Cybersecurity Engineering, Politeknik Negeri Cilacap, Cilacap, Indonesia

Received:
December 12, 2025
Revised:
March 10, 2026
Accepted:
April 1, 2026
Published:
April 12, 2026

Corresponding Author:
Author Name*:
Ratih Hafsarah Maharrani
Email*:
ratih.hafsarah@pnc.ac.id

DOI:
10.63158/journalisi.v8i2.1519

Abstract. Computer networks are vital to modern organizations, yet growing digital dependence has increased both the frequency and complexity of cyberattacks. To address this challenge, this study proposes an interpretable cyberattack detection framework that combines rule-based reasoning with machine learning through a hybrid Certainty Factor (CF)–XGBoost model. The framework integrates CF confidence scores and XGBoost probability outputs within a meta-learning classifier, enabling strong predictive performance while preserving explainability. Experiments conducted on the TON_IoT dataset using an 80:20 stratified split demonstrate that XGBoost achieved the highest accuracy at 99.61%, followed closely by the hybrid model at 99.42%, whereas the standalone CF model reached 76.31%. Although the hybrid approach produced a slightly lower accuracy than XGBoost alone, it substantially enhanced interpretability by connecting predictions to explicit rule-based reasoning. This makes the proposed framework especially suitable for Security Operations Center (SOC) environments, where transparent decision-making is essential. Overall, the findings suggest that the hybrid CF–XGBoost model offers a practical and explainable solution for cyberattack detection, though further validation on more diverse datasets is necessary before real-world deployment.

Keywords: intrusion detection, TON_IoT, Certainty Factor, XGBoost, explainable cyberattack detection

1. INTRODUCTION

Cyberattacks refer to malicious activities targeting information systems and computer networks with the intention of gaining unauthorized access, disrupting operations, or causing system damage [1]. In recent years, the rapid advancement of information technology has increased the dependence of various sectors, such as government, business, education, and public services, making computer networks fundamental to maintaining the continuity of modern systems. However, this growing connectivity, both internal and external has also intensified the risk, frequency, and sophistication of cyberattacks. These threats can seriously affect the confidentiality, integrity, and availability of information systems, highlighting the need for detection mechanisms that are not only accurate but also reliable and timely. Recent studies on intrusion detection systems also emphasize that modern cyberattack detection must address not only predictive performance but also on explainability, particularly in dynamic IoT and Industry 4.0 environments where network behavior is highly diverse and constantly evolving [2]

Efforts to strengthen network security increasingly incorporate artificial intelligence-based detection models. One established approach is the use of expert knowledge-based systems, which rely on structured rules and domain knowledge to support decision-making [3]. In cybersecurity, such systems can provide early indications of attacks based on observed patterns in network traffic. The Certainty Factor (CF) method is commonly used to represent the degree of confidence in a given rule or hypothesis [4]. An important advantage of CF is its ability to produce decisions that can be clearly traced and explained through underlying rules, making it particularly suitable for applications where interpretability is essential.

Despite this advantage, the Certainty Factor method faces limitations when applied to modern network data, which is often complex, high-dimensional, and continuously evolving. Because it depends on predefined rules, CF is less effective in adapting to previously unseen attack patterns. This limitation suggests the need for a complementary approach that can learn directly from data. To address this, the present study integrates CF with the Extreme Gradient Boosting (XGBoost) algorithm. XGBoost is a powerful boosting-based machine learning method that improves model performance iteratively while incorporating regularization to reduce overfitting [5]. In this context, CF contributes

interpretable reasoning, whereas XGBoost enhances detection performance by capturing complex patterns within large-scale data. Previous studies have shown that XGBoost consistently delivers strong results in intrusion detection tasks, although its limited transparency remains a concern for practical deployment [6] A number of studies have explored machine learning techniques for cyberattack detection. Random Forest, for example, has demonstrated strong classification performance; however, it is often regarded as a black-box model due to its limited interpretability. Similarly, network forensic approaches are effective for post-incident analysis but remain inherently reactive, as they are applied after an attack has occurred. Meanwhile, boosting-based models such as XGBoost have achieved high predictive performance across various intrusion detection benchmarks. Even so, many of these approaches still lack sufficient transparency, which can limit their usefulness in real-world security decision-making [2].

Taken together, existing studies point to a clear trade-off between predictive performance and interpretability. At the same time, forensic methods tend to be reactive, while rule-based approaches struggle to adapt to evolving threats. This observation underscores the need for a more balanced detection framework. Accordingly, the main research gap addressed in this study is the limited availability of cyberattack detection models that effectively combine interpretable rule-based reasoning with high-performance machine learning within a unified framework, particularly when evaluated on realistic IoT-based datasets. The TON_IoT dataset is specifically designed to support the evaluation of cybersecurity models in IoT and IIoT environments, making it highly relevant for this purpose [7], [8], [9], [10]. Unlike prior approaches that rely solely on rule-based reasoning or purely data-driven models, this study introduces a hybrid CF–XGBoost approach that brings both perspectives together. A key novelty of this work is the use of a meta-learning framework, in which CF-derived scores and XGBoost prediction probabilities are combined as input features for a secondary XGBoost classifier, forming an integrated detection pipeline [6].

This study makes three main contributions. First, it proposes an explainable hybrid detection model that integrates rule-based confidence reasoning with machine learning classification. Second, it provides a comparative evaluation of CF, XGBoost, and the hybrid CF–XGBoost model using the TON_IoT dataset. Third, it assesses model performance using confusion matrix metrics, ROC analysis, and feature importance analysis to evaluate

both predictive accuracy and interpretability. Overall, this study aims to develop a cyberattack detection approach that achieves a meaningful balance between strong predictive performance and transparent decision-making.

2. METHODS

This study proposes a hybrid cyberattack detection framework that integrates rule-based reasoning using the Certainty Factor (CF) method with machine learning classification through Extreme Gradient Boosting (XGBoost), as illustrated in Fig. 1. The proposed framework consists of six main stages: dataset preparation, data preprocessing, Certainty Factor knowledge-based construction, CF score computation, XGBoost and hybrid model training, and final performance evaluation. The study compares three detection models: (1) a standalone CF-based rule system, (2) a standalone XGBoost classifier, and (3) a hybrid CF–XGBoost model in which the the CF score and XGBoost predicted probability are used as inputs to a second-stage classifier, rather than being combined through a simple voting scheme.

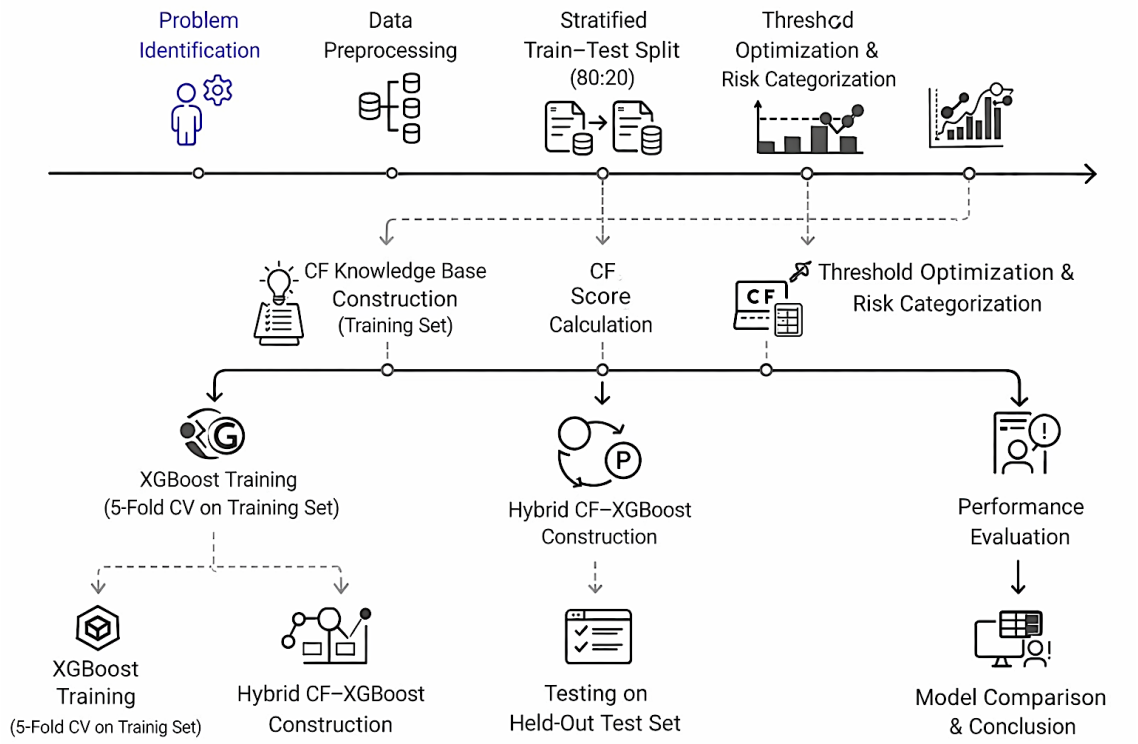

Figure 1. Research Methodology Flowchart

2.1. Dataset

Experiments were conducted using the TON_IoT dataset developed by Moustafa et al. at the University of New South Wales [9]. The dataset contains 211,043 network traffic records consisting of 50,000 normal samples and 161,043 attack samples, indicating a significant class imbalance. Because of this imbalance, evaluation was not limited to accuracy alone; balanced accuracy and Matthews Correlation Coefficient (MCC) were also included to provide a more reliable assessment. No resampling techniques (such as oversampling or undersampling) were applied, and no class-weighting strategies were used during model training. The original data distribution was fully preserved to avoid introducing additional bias. Recent studies also highlight that class imbalance can affect detection performance, particularly for minority classes, and should therefore be considered explicitly in evaluation [11].

2.2. Data Preprocessing

Before model training, the dataset underwent preprocessing to improve data quality and analytical consistency. Irrelevant textual attributes, such as IP-, DNS-, SSL-, and HTTP-related fields not directly used in detection, were removed. Numerical features were converted into numeric format, invalid values were handled, and highly skewed variables were log-transformed when necessary. Missing numeric values were replaced with zero, ensuring consistent feature representation without introducing additional assumptions through imputation. Categorical features were standardized and encoded before model training. In addition, two derived features, bytes_ratio and packet_rate, were generated to enrich traffic representation.

2.3. Training–Testing Design

The dataset was split into 80% training data and 20% testing data using stratified sampling. All rule generation, threshold definition, and model training were performed on the training set only. Five-fold cross-validation was applied during XGBoost development, while all final confusion matrices and performance metrics were computed on the held-out test set, ensuring that no information from the test data influenced model construction.

2.4. Certainty Factor Knowledge Base Construction

The Certainty Factor (CF) method was employed to provide interpretable, rule-based reasoning. Moving away from traditional manual rule definition, the CF knowledge base in this study was generated automatically from the training set. Candidate rules were extracted from selected categorical features (*conn_state, proto, service, http_status_code, weird_notice, http_method, ssl_resumed, and http_uri*) and numerical features (*duration, src_bytes, dst_bytes, dns_qtype, bytes_ratio, and packet_rate*).

The Measure of Belief (MB) and Measure of Disbelief (MD) values were calculated based on conditional probability distribution [12] formulated as shown in Equation 1 to 3.

$$MB = \frac{max\,(0, P(Attack|Feature) - P(Normal|Feature))}{P(Attack|Feature) + P(Normal|Feature) + \epsilon} \quad (1)$$

$$MD = \frac{max\,(0, P(Normal|Feature) - P(Attack|Feature))}{P(Attack|Feature) + P(Normal|Feature) + \varepsilon} \quad (2)$$

with the rule certainty defined as:

$$CF = MB - MD \quad (3)$$

To ensure robustness, categorical rules were strictly filtered using a minimum support threshold of 0.05 (5%) and a minimum confidence threshold of 0.60 (60%), coupled with an internal condition that either the MB or MD must exceed 0.80. This rigorous filtering process resulted in the generation of 16 highly reliable rules derived entirely from the training dataset, ensuring that the knowledge base remains both compact and statistically significant.

2.5. CF Score Computation

For each traffic record, activated rules were combined iteratively to obtain a total CF value. The final CF value was then transformed using a modified sigmoid function to normalize the score into the range [0, 1], thereby improving stability for downstream classification.

2.6. Threshold Definition and Risk Categorization

The continuous CF score was utilized for both binary classification and risk interpretation. The binary decision threshold was determined through precision–recall curve optimization on the training data. For interpretive analysis, CF scores were clustered into three groups using the K-Means algorithm and then mapped to low-, medium-, and high-risk categories through adaptive thresholds.

2.7. XGBoost Classification Model

XGBoost was used as the machine learning classifier because of its strong predictive capability and regularization mechanism [13], [14]. The model was trained only on the training subset after excluding labels, CF-related outputs, and removed features. Categorical variables were encoded using label encoding, and five-fold cross-validation was applied during training-set development [15].. The main parameters were 100 trees, maximum depth 3, learning rate 0.1, subsample 0.8, colsample_bytree 0.8, gamma 0.1, reg_alpha 0.5, and reg_lambda 1.0.

2.8. Hybrid CF–XGBoost Integration

The core novelty of this study lies in the proposed hybrid CF–XGBoost framework. Unlike conventional hybrid approaches that simply append additional variables to the original high-dimensional feature space, this method employs a compact meta-learning strategy. Specifically, the second-stage classifier utilizes only two inputs: the predicted probability from the XGBoost model ($P_{XGBoost}$) and the Certainty Factor score (Score_{CF}) [16], defined as shown in Equation 4.

$$X_{hybrid} = \left[X_{Original}, CF_{score}\right] \tag{4}$$

These two values were used to train a second-stage XGBoost classifier, allowing statistical prediction confidence and rule-based confidence to be combined in one decision process [17].

2.9. Performance Evaluation

Model performance was comprehensively evaluated using accuracy, precision, recall, F1-score, balanced accuracy, ROC curves, and Area Under the Curve (AUC). All final confusion matrices and performance metrics were computed exclusively on the held-out test set,

ensuring a fair and unbiased comparison among the standalone CF, standalone XGBoost, and the hybrid CF–XGBoost models. [18], [19], [20].

3. RESULTS AND DISCUSSION

3.1 Dataset Characteristics

The dataset used in this study consists of 211,043 network traffic records that underwent comprehensive preprocessing. The data were divided into two classes: 50,000 normal instances and 161,043 attack instances, indicating a significant class imbalance (23.7% normal vs. 76.3% attack). This imbalance accurately reflects realistic intrusion detection conditions, where malicious traffic often dominates certain vulnerable network environments and poses significant challenges for classification models.

3.2 Certainty Factor (CF) Knowledge Base Construction

The proposed Certainty Factor (CF) rule-based system successfully constructed an automatic knowledge base consisting of 16 symptom-based rules extracted from the training dataset. Unlike conventional expert systems that rely entirely on manually defined heuristics, this study adopts a data-driven rule formulation, where confidence values are derived from the statistical distribution of features across attack and normal traffic [21]. Measure of Belief (MB) and Measure of Disbelief (MD) values were computed using conditional probability analysis as described in the methodology section. The resulting CF confidence score reflects the strength of association between network traffic characteristics and the corresponding class label. This probabilistic approach allows the CF model to perform inference based on real network evidence, making it more adaptable to large-scale cybersecurity datasets while maintaining explainability [5]

Based on feature distribution analysis, the system generated interpretable rules representing relationships between network attributes and cyberattack behavior. For instance, the condition conn_state = 'sf' produced a high belief value for the normal class, as it indicates successful connection termination, which is dominant in benign traffic. Similarly, the rule duration > Q95 captures unusually long connection durations, which may indicate anomalous activity such as slow-rate attacks or abnormal data transfers. The condition src_bytes < Q5 reflects extremely low source byte counts, often associated with scanning or abnormal communication patterns. These automatically extracted rules

demonstrate that the CF model remains interpretable, as each prediction can be traced back to specific traffic conditions, while still being adaptive to real-world network distributions.

3.3 Evaluation of the Certainty Factor Model

The standalone CF model achieved an overall accuracy of 76.31%. Adaptive decision thresholds of 0.225 (low-risk) and 0.788 (medium-risk) were applied to separate attack from normal traffic. These results confirm that while CF provides an interpretable baseline risk indicator, its predictive performance remains highly limited when handling complex and diverse cyberattack patterns without machine learning assistance [22]

As illustrated in Fig. 2, the CF model classified all traffic as attacks, resulting in 32,209 true positives but missing all normal samples (0 true negatives). Consequently, 10,000 normal instances were incorrectly classified as attacks (false positives), while 0 attacks were missed (false negatives). This outcome demonstrates that CF achieves maximum sensitivity toward malicious activity, yet its ability to discriminate normal traffic is entirely lacking in this standalone configuration. The 100% false positive rate for benign traffic would lead to severe alert fatigue in operational environments. Therefore, CF is strictly more suitable as a supportive risk indicator within a hybrid detection framework rather than as a standalone classifier. This behavior indicates that the generated CF rules are biased toward attack-dominant patterns due to the imbalanced nature of the dataset.

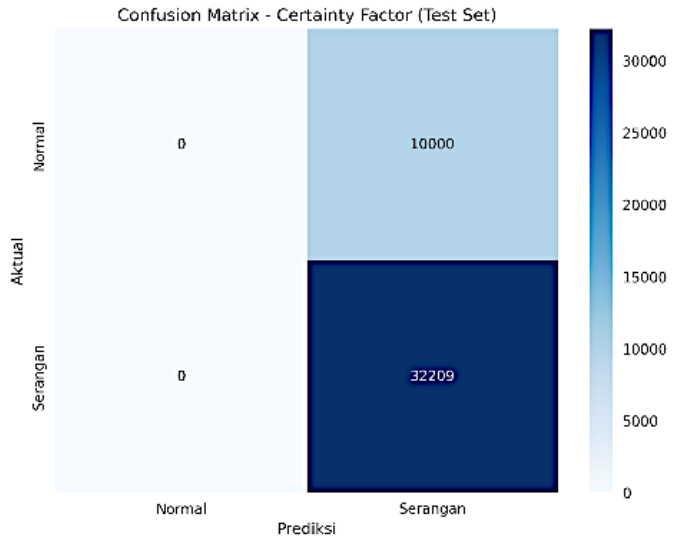

Figure 2. Confusion matrix of the Certainty Factor model on the test dataset.

The CF approach produced an extremely high recall for the attack class (100%), highlighting its reliability in capturing potentially harmful traffic. Its primary advantage lies in explainability, as predictions can be directly traced back to activated rules, and in computational efficiency, since it does not require extensive training data. Nevertheless, normal-class precision remained unacceptably low (24%), confirming that all benign connections were misclassified as attacks. The balanced accuracy value further indicates a severe bias toward the majority attack class. These characteristics strongly suggest that CF is most effective when integrated with more precise machine learning classifiers to reduce false alarms.

The risk-level classification results, shown in Fig. 3, successfully partitioned the CF scores into three distinct clusters: low risk (CF < 0.225), medium risk (0.225 ≤ CF ≤ 0.788), and high risk (CF > 0.788). This distribution shows that the K-Means algorithm was able to separate the network flows into interpretable risk groups. In practical terms, this means that the CF model is not only useful for binary attack indication but also for highlighting which traffic deserves closer attention first. Such information can support early warning systems and help security analysts prioritize mitigation efforts more effectively.

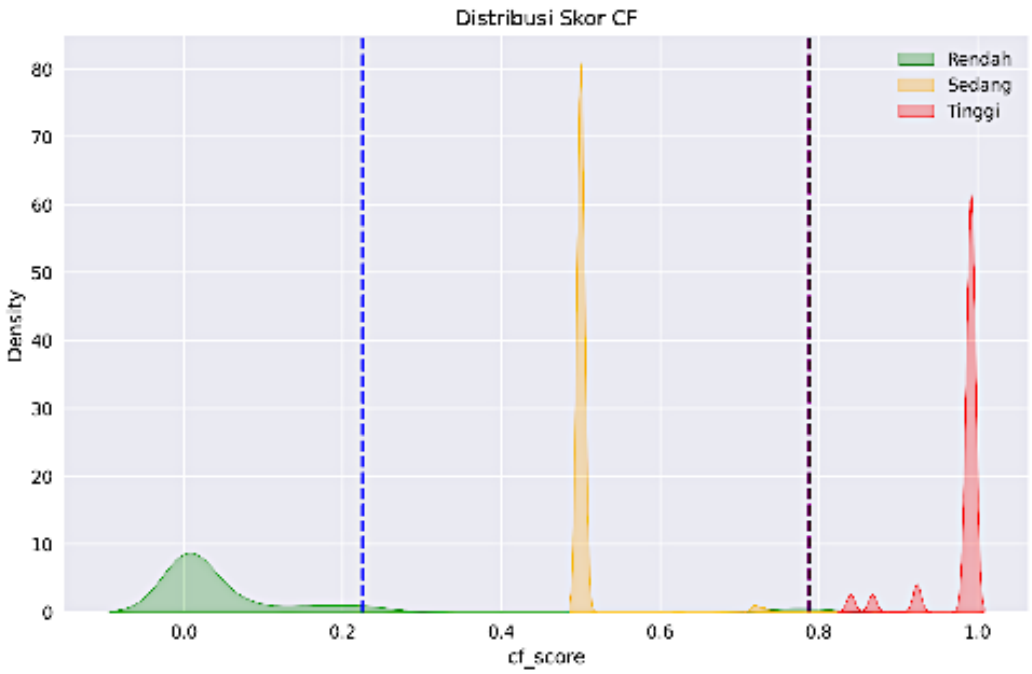

Figure 3. Distribution of risk levels in network traffic classification results based on Certainty Factor scores.

3.4 Evaluation of the XGBoost Model

The XGBoost classifier achieved the highest performance among all models, with an accuracy of 99.61% and a balanced accuracy of 99.45%. The 5-fold cross-validation score of 99.51% ± 0.03% demonstrates consistent performance across different data subsets, confirming the robustness of XGBoost for cyberattack detection tasks.

As presented in Fig. 4, XGBoost correctly classified 9,916 normal samples and 32,128 attack samples, producing only 84 false positives and 81 false negatives. These minimal error rates (0.8% misclassification for normal traffic and 0.25% for attacks) highlight the model's strong generalization capability, balanced detection across both classes, and significantly reduced false alarm generation, making it highly suitable for real-world intrusion detection.

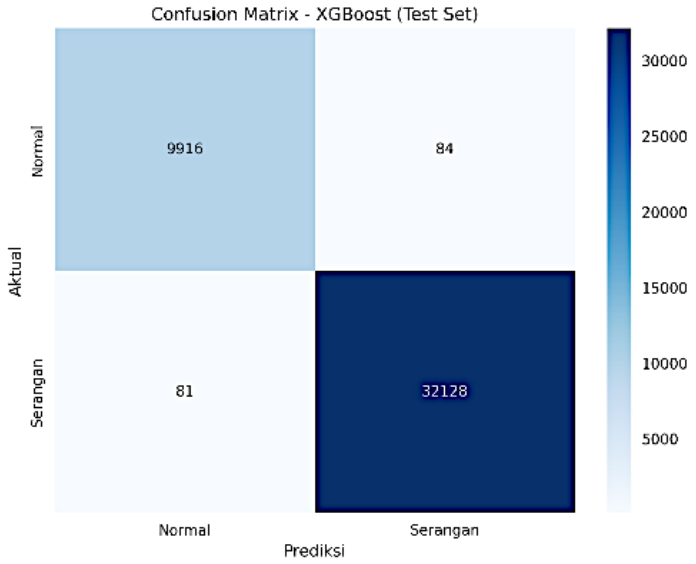

Figure 4. Confusion matrix of the XGBoost model on the test dataset.

Feature importance results in Fig. 5 indicate that proto, dns_rcode, and dns_AA contributed most significantly to the XGBoost classification decisions, with importance scores of 0.160, 0.120, and 0.101, respectively. Unlike models that rely too heavily on a single dominant feature, the highest importance value in this study remains relatively low. This suggests that the classifier learns from a combination of network attributes rather than depending on only one signal. Such a pattern is crucial because it usually reflects a more stable model that is less likely to become unreliable when traffic

characteristics change in unseen environments. This balanced contribution across features strongly supports the model's robustness.

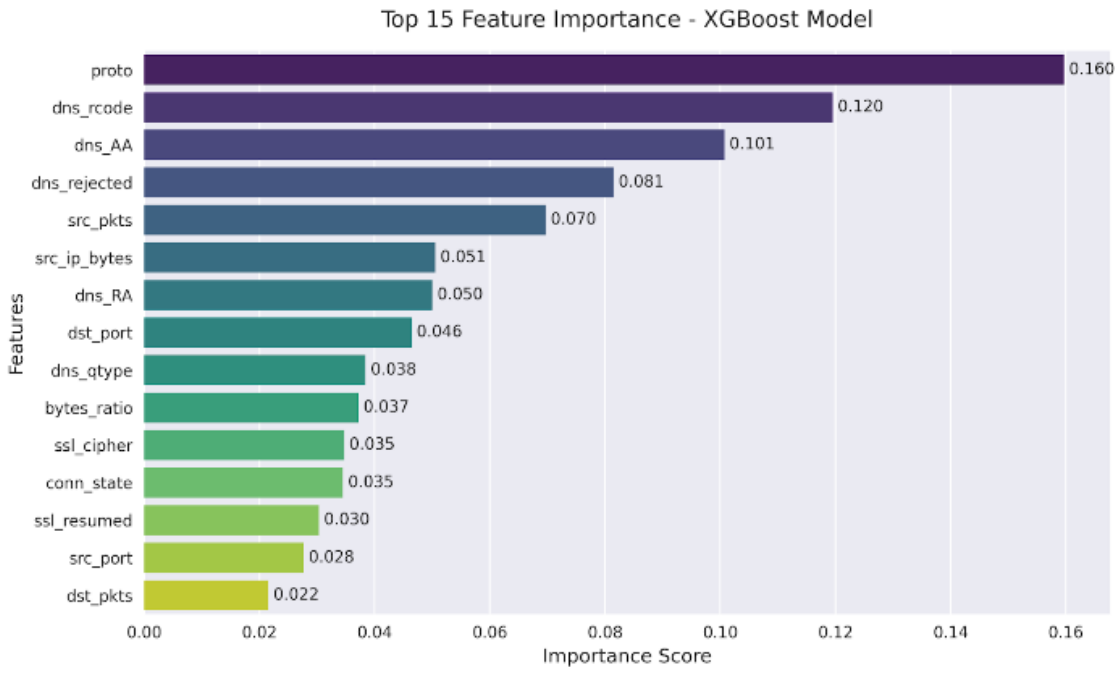

Figure 5. Feature importance in the XGBoost model, highlighting proto and dns_rcode as the top contributing network attributes.

3.5 Evaluation of the Hybrid CF–XGBoost Model

The hybrid CF–XGBoost model achieved an accuracy of 99.42% and a balanced accuracy of 99.09%. Cross-validation performance reached 99.43%, indicating strong stability and resistance to overfitting. This hybrid approach effectively combines CF-based interpretability with XGBoost's predictive strength, producing a well-balanced detection model.

As shown in Fig. 6, the hybrid model correctly classified 9,855 normal samples and 32,119 attack samples, with only 145 false positives and 90 false negatives. These findings confirm that incorporating CF confidence reasoning into XGBoost provides additional risk interpretability while maintaining exceptionally high detection specificity [23]

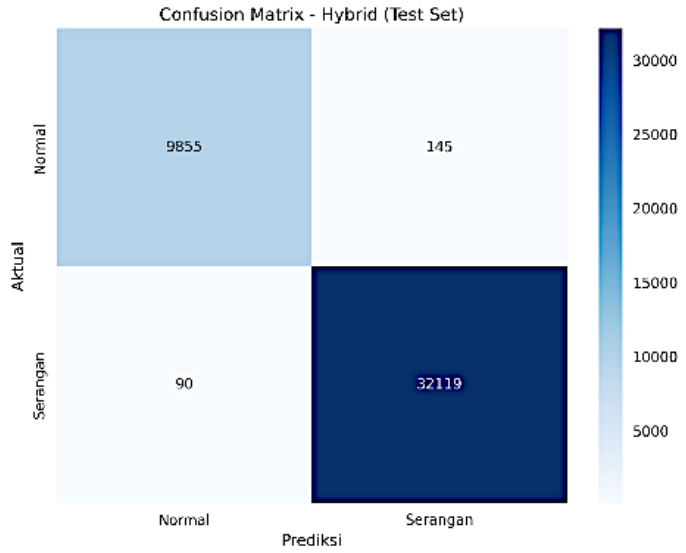

Figure 6. Confusion matrix of the hybrid CF–XGBoost model on the test dataset

As shown in Figure 6 the evaluation results, the hybrid model yielded a slightly lower accuracy (99.42%) and balanced accuracy (99.09%) compared to the standalone XGBoost model (99.61% and 99.45%, respectively). This slight performance drop is expected, as the inclusion of the rigid, rule-based CF scores acts as a mild constraint on the highly flexible gradient boosting process. However, the operational value of the hybrid approach heavily outweighs this minor 0.19% decrease in raw accuracy. In real-world Security Operations Centers (SOCs), a 'black-box' model like standalone XGBoost often leads to alert fatigue and prolonged investigation times because analysts cannot trace why an alert was generated. The hybrid model bridges this gap by inherently tying its predictions to human-readable rule activations (CF scores), facilitating significantly faster, more transparent, and more confident incident response.

3.6 Comparative Analysis of All Models

Comparative evaluation shows that the CF model achieved the lowest performance (76.31% accuracy) due to its massive false positive rate. In contrast, XGBoost produced the best overall results (99.61%), demonstrating superior consistency. The hybrid CF–XGBoost model achieved an accuracy of 99.42%, closely approaching the standalone XGBoost performance while simultaneously offering the crucial benefit of rule-based interpretability.

The ROC comparison in Fig. 7 highlights clear performance differences. The CF model obtained an AUC of 0.61, indicating that its standalone discrimination ability is only slightly better than random guessing. This confirms that although CF is highly sensitive to attack-related evidence, it is weak in separating normal and malicious traffic when used alone. By contrast, both XGBoost and the hybrid model achieved an AUC of 1.00, showing excellent separability across decision thresholds. These findings suggest that the hybrid framework successfully preserves the predictive strength of XGBoost while natively integrating CF-based reasoning. This result may indicate a high degree of class separability within the TON_IoT dataset or potential dataset-specific bias, rather than reflecting perfect generalization in real-world environments.

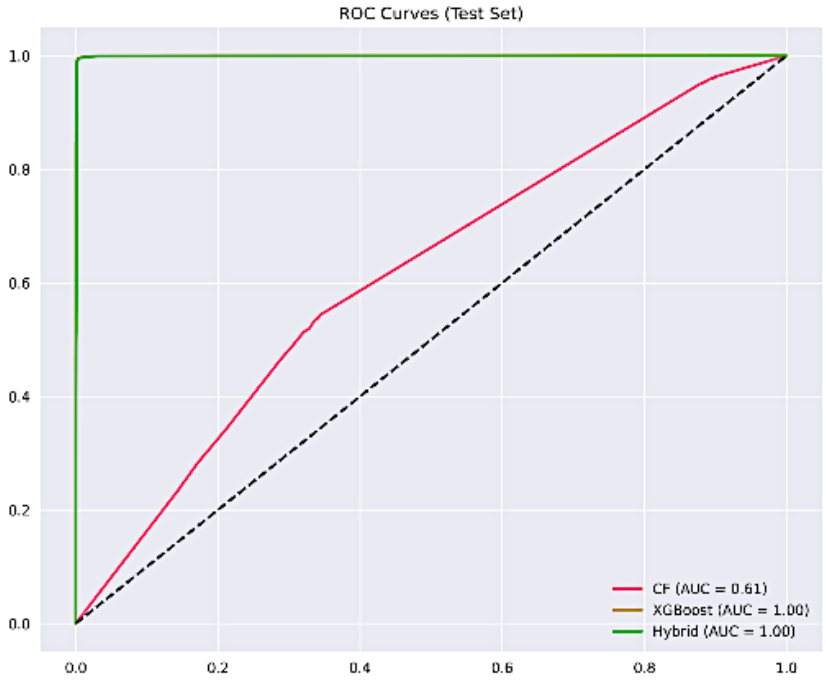

Figure 7. ROC curves comparing the CF, XGBoost, and Hybrid CF–XGBoost models, showing AUC values of 0.61, 1.00, and 1.00, respectively.

3.7 Discussion

The overall findings suggest that the three models serve distinct yet complementary roles in cyberattack detection. The standalone CF model offers strong interpretability and very high sensitivity to malicious traffic, making it useful for identifying potentially harmful events from a rule-based perspective. However, its limited ability to correctly recognize normal traffic leads to an excessively high false positive rate, which reduces

its practicality in operational settings. This outcome indicates that rule-based reasoning alone is not sufficient for handling modern network traffic, particularly when the data are complex and imbalanced. In contrast, XGBoost delivered the highest predictive performance and the most consistent classification results, highlighting its suitability for automated detection tasks where accuracy and robustness are the primary priorities. The proposed hybrid CF–XGBoost model provides an effective compromise by maintaining performance close to that of standalone XGBoost while preserving the interpretability advantages of CF. As a result, the hybrid approach supports a more transparent and explainable decision-making process, which is especially valuable in operational cybersecurity contexts where analysts must justify alerts and response actions.

Although the near-perfect performance metrics obtained in this study, including accuracy above 99% and an AUC of 1.00, demonstrate the effectiveness of the proposed framework, these results should be interpreted with caution. Such exceptionally high values may partly reflect the structured and well-defined nature of benchmark datasets rather than the full complexity of real-world network environments. In particular, an AUC of 1.00 suggests almost perfect class separability, a condition that is rarely sustained in live operational settings where traffic patterns are more diverse, noisy, and continuously evolving. A major limitation of this study is the use of a single static dataset, TON_IoT, which may contain dataset-specific characteristics, inherent biases, or artifacts that simplify the classification problem. In real deployments, intrusion detection systems must operate under concept drift, changing attack strategies, encrypted traffic, and previously unseen zero-day threats. Therefore, the current results should be viewed as a strong baseline that demonstrates the promise of the proposed framework, rather than as definitive evidence of real-world performance.

The performance of the proposed hybrid CF–XGBoost framework is broadly consistent with recent intrusion detection studies that also use the TON_IoT dataset, while offering a notable advantage in terms of interpretability. Recent studies based on hybrid deep learning architectures frequently report accuracy levels above 99%; however, these models often involve substantial computational complexity and do not provide inherent explainability [19]. In comparison, traditional ensemble methods such as Random Forest also achieve strong classification performance and high AUC values, but they rely on complex non-linear decision boundaries that are difficult to interpret directly, often

requiring post-hoc explanation techniques such as SHAP to support analysis and trustworthiness [24], [2]. This distinction is important in cybersecurity operations, where model transparency can influence analyst confidence, incident prioritization, and response efficiency. By integrating CF-derived risk scores directly into the XGBoost meta-classifier, the proposed method embeds explainability into the detection pipeline itself rather than treating it as an additional interpretive layer.

This design choice allows the hybrid framework to achieve a competitive accuracy of 99.42% while still providing transparent rule-based reasoning for its decisions. Although its performance is slightly lower than that of standalone XGBoost, the difference is marginal when considered against the practical benefit of improved interpretability. In many real-world SOC environments, a small reduction in predictive performance may be acceptable if it is accompanied by a significant gain in explainability, especially when analysts need to investigate alerts, validate suspicious behavior, and communicate findings to stakeholders. The results therefore demonstrate that interpretability does not necessarily need to be sacrificed to achieve high detection performance. Instead, it can be incorporated directly into the model architecture, reducing dependence on post-hoc explanation methods and improving the usability of the system in operational cybersecurity workflows.

Future work should extend the evaluation of the proposed framework across more diverse and widely used intrusion detection datasets, such as CICIDS2017 and UNSW-NB15, in order to assess its generalizability under different traffic distributions and attack scenarios. Additional comparisons with baseline classifiers, including Support Vector Machines (SVM), Random Forest, and deep learning-based models, would also help position the framework more comprehensively within the current intrusion detection literature. Beyond cross-dataset validation, further investigation into real-time performance, resilience to concept drift, and effectiveness against zero-day attacks would strengthen the case for deployment in realistic environments. Overall, the present findings are comparable to recent TON_IoT-based IDS studies reporting accuracy above 99%, while offering the additional benefit of improved interpretability, which remains a critical requirement for trustworthy and practical cyberattack detection systems.

4. CONCLUSION

This study proposed a hybrid CF–XGBoost framework for cyberattack detection that integrates rule-based reasoning with machine learning. Experimental results show that XGBoost achieved the highest predictive performance (99.61% accuracy and 99.45% balanced accuracy), while the proposed hybrid model delivered highly competitive results (99.42% accuracy and 99.09% balanced accuracy) with the added advantage of interpretability through CF-based reasoning. The findings demonstrate that high detection performance can be achieved without sacrificing transparency. While the standalone CF model provides strong interpretability, its high false positive rate limits its standalone applicability, reinforcing the importance of hybrid approaches. However, the near-perfect performance metrics obtained in this study are closely associated with the characteristics of the TON_IoT dataset. Therefore, future work should focus on evaluating the proposed framework on diverse and real-world datasets to assess its generalizability and robustness in dynamic network environments.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Politeknik Negeri Cilacap through the Center for Research and Community Service (P3M) for the support and facilitation provided, which enabled this research to be conducted and completed successfully.

REFERENCES

[1] Y. Li and Q. Liu, "A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments," *Energy Reports*, vol. 7, 2021, doi: 10.1016/j.egyr.2021.08.126.

[2] V. Z. Mohale and I. C. Obagbuwa, "Evaluating machine learning-based intrusion detection systems with explainable AI: enhancing transparency and interpretability," *Front. Comput. Sci.*, vol. 7, 2025, doi: 10.3389/fcomp.2025.1520741.

[3] X. J. Tan, W. L. Cheor, K. S. Yeo, and W. Z. Leow, "Expert systems in oil palm precision agriculture: A decade systematic review," 2022. doi: 10.1016/j.jksuci.2022.02.006.

[4] Sumiati, H. Saragih, T. K. A. Rahman, and A. Triayudi, "Expert system for heart disease based on electrocardiogram data using certainty factor with multiple rule," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 1, 2021, doi: 10.11591/ijai.v10.i1.pp43-50.

[5] L. Theodorakopoulos, A. Theodoropoulou, A. Tsimakis, and C. Halkiopoulos, "Big Data-Driven Distributed Machine Learning for Scalable Credit Card Fraud Detection Using PySpark, XGBoost, and CatBoost," *Electronics (Switzerland)*, vol. 14, no. 9, 2025, doi: 10.3390/electronics14091754.

[6] Y. Hu, K. Xiao, L. Luo, and L. Chen, "An XGBoost-Based Intrusion Detection Framework with Interpretability Analysis for IoT Networks," *Applied Sciences*, vol. 16, no. 2, 2026, doi: 10.3390/app16020980.

[7] N. Moustafa, "New Generations of Internet of Things Datasets for Cybersecurity Applications based Machine Learning: TON_IoT Datasets," *eResearch Australia Asia 2019*, no. October, 2019.

[8] T. M. Booij, I. Chiscop, E. Meeuwissen, N. Moustafa, and F. T. H. D. Hartog, "ToN_IoT: The Role of Heterogeneity and the Need for Standardization of Features and Attack Types in IoT Network Intrusion Data Sets," *IEEE Internet Things J.*, vol. 9, no. 1, 2022, doi: 10.1109/JIOT.2021.3085194.

[9] N. Moustafa, "A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets," *Sustain. Cities Soc.*, vol. 72, 2021, doi: 10.1016/j.scs.2021.102994.

[10] A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and A. Anwar, "TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems," *IEEE Access*, vol. 8, pp. 165130–165150, 2020, doi: 10.1109/ACCESS.2020.3022862.

[11] V. Shanmugam, R. Razavi-Far, and E. Hallaji, "Addressing Class Imbalance in Intrusion Detection: A Comprehensive Evaluation of Machine Learning Approaches," *Electronics (Basel)*, vol. 14, no. 1, p. 69, Dec. 2024, doi: 10.3390/electronics14010069.

[12] O. Galal, A. Nasr, and L. Rizkallah, "A Rule Learning Approach For Building An Expert System To Detect Network Intrusions," *International Journal of Intelligent Computing and Information Sciences*, vol. 23, no. 1, pp. 106–114, Mar. 2023, doi: 10.21608/ijicis.2023.167424.1223.

[13] S. Thongsuwan, S. Jaiyen, A. Padcharoen, and P. Agarwal, "ConvXGB: A new deep learning model for classification problems based on CNN and XGBoost," *Nuclear Engineering and Technology*, vol. 53, no. 2, 2021, doi: 10.1016/j.net.2020.04.008.

[14] S. M. Nzuva, L. Nder, and T. Mwalili, "A novel bagging- XGBoost ensemble model for attaining high accuracy and computational efficiency in network intrusion detection," *E3S Web of Conferences*, vol. 501, p. 01007, Mar. 2024, doi: 10.1051/e3sconf/202450101007.

[15] J. Vitorino, R. Andrade, I. Praça, O. Sousa, and E. Maia, "A Comparative Analysis of Machine Learning Techniques for IoT Intrusion Detection," 2022, pp. 191–207. doi: 10.1007/978-3-031-08147-7_13.

[16] N. Saini, V. Bhat Kasaragod, K. Prakasha, and A. K. Das, "A hybrid ensemble machine learning model for detecting APT attacks based on network behavior anomaly detection," *Concurr. Comput.*, vol. 35, no. 28, Dec. 2023, doi: 10.1002/cpe.7865.

[17] A. M. Aburbeian, M. Fernández-Veiga, and A. Hasasneh, "Improving Remote Access Trojans Detection: A Comprehensive Approach Using Machine Learning and Hybrid Feature Engineering," *AI*, vol. 6, no. 9, p. 237, Sep. 2025, doi: 10.3390/ai6090237.

[18] D. Chicco, N. Tötsch, and G. Jurman, "The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation," *BioData Min.*, vol. 14, no. 1, p. 13, Feb. 2021, doi: 10.1186/s13040-021-00244-z.

[19] Md. N. Sarwar, Md. S. Arman, T. Bhuiyan, and F. B. Rafiq, "Optimizing Intrusion Detection with Hybrid Deep Learning Models and Data Balancing Techniques," in *2025 IEEE 4th International Conference on AI in Cybersecurity (ICAIC)*, IEEE, Feb. 2025, pp. 1–6. doi: 10.1109/ICAIC63015.2025.10849340.

[20] J. N. Mandrekar, "Receiver Operating Characteristic Curve in Diagnostic Test Assessment," *Journal of Thoracic Oncology*, vol. 5, no. 9, pp. 1315–1316, Sep. 2010, doi: 10.1097/JTO.0b013e3181ec173d.

[21] H. Liu and B. Lang, "Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey," *Applied Sciences*, vol. 9, no. 20, p. 4396, Oct. 2019, doi: 10.3390/app9204396.

[22] A. Kukliansky, M. Orescanin, C. Bollmann, and T. Huffmire, "Network Anomaly Detection Using Quantum Neural Networks on Noisy Quantum Computers," *IEEE Transactions on Quantum Engineering*, vol. 5, 2024, doi: 10.1109/TQE.2024.3359574.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[23] A. Haque and H. Soliman, "A Transformer-Based Autoencoder with Isolation Forest and XGBoost for Malfunction and Intrusion Detection in Wireless Sensor Networks for Forest Fire Prediction," *Future Internet*, vol. 17, no. 4, 2025, doi: 10.3390/fi17040164.

[24] S. M. Nzuva, L. Nder, and T. Mwalili, "A novel bagging- XGBoost ensemble model for attaining high accuracy and computational efficiency in network intrusion detection," *E3S Web of Conferences*, vol. 501, p. 01007, Mar. 2024, doi: 10.1051/e3sconf/202450101007.